Genomics 97 (2011) 173–185

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

# Rapid divergence of repetitive DNAs in *Brassica* relatives

Dal-Hoe Koo [a,d,1,2], Chang Pyo Hong [b,f,1,2], Jacqueline Batley [c], Yong Suk Chung [d,2], David Edwards [e], Jae-Wook Bang [a], Yoonkang Hur [a,*], Yong Pyo Lim [b,*]

[a] Department of Biological Science, College of Biological Science and Biotechnology, Chungnam National University, Daejeon 305-764, Republic of Korea
[b] Department of Horticulture, College of Agriculture and Life Science, Chungnam National University, Daejeon 305-764, Republic of Korea
[c] ARC Centre of Excellence for Integrative Legume Research and School of Land, Crop, and Food Sciences, University of Queensland, Brisbane, Qld 4072, Australia
[d] Department of Horticulture, University of Wisconsin-Madison, Madison, WI 53706, USA
[e] Australian Centre for Plant Functional Genomics, Institute for Molecular Biosciences and School of Land, Crop, and Food Sciences, University of Queensland, Brisbane, Qld 4072, Australia
[f] Department of Life Science, and BK21 Program of Bio-Molecular Function, Pohang University of Science and Technology, Pohang 790-784, Republic of Korea

ARTICLE INFO

*Article history:*
Received 29 July 2010
Accepted 2 December 2010
Available online 13 December 2010

*Keywords:*
*Brassica* species
Centromere
Evolution
Subtelomere
Telomere

ABSTRACT

Centromeric, subtelomeric, and telomeric repetitive DNAs were characterized in *Brassica* species and the related *Raphanus sativus* and *Arabidopsis thaliana*. In general, rapid divergence of the repeats was found. The centromeric tandem satellite repeats were differentially distributed in the species studied, suggesting that centromeric repeats have diverged during the evolution of the A/C and B genome lineages. Sequence analysis of centromeric repeats suggested rapid evolution. Pericentromere-associated retrotransposons were identified and showed divergence during the evolution of the lineages as centromeric repeats. A novel subtelomeric tandem repeat from *B. nigra* was found to be conserved across the diploid *Brassica* genomes; however, this sequence was not identified in the related species. In contrast to previous studies, interstitial telomere-like repeats were identified in the pericentromeres of *Brassica* chromosomes, and these repeats may be associated with genomic stability. These results provide insight into genome evolution during polyploidization in *Brassica* and divergence within the Brassicaceae.



## 1. Introduction

Advances in genome studies have shown that eukaryotic genomes contain a significant proportion of repetitive DNA sequences. In plant genomes, repetitive sequences are widespread and responsible for the heterogeneity in genome size and composition [1]. For example, the maize genome contains approximately 2500 Mbp of DNA, 58–80% of which is estimated to be repetitive [2]. Approximately 90% of the 4000 Mbp of the *Aegilops tauschii* genome, which contributes the D genome of bread wheat, is repetitive [3], and even the comparatively small genome of the model plant *Arabidopsis thaliana* (146 Mbp) contains ~25% repeat sequence (reviewed in Thompson et al. [4]).

Repetitive DNA sequences include tandem repeats, dispersed repeats, retroelements, and rDNAs. Tandemly arrayed repetitive DNA sequences, known as satellite DNA, are ubiquitous and are representative constituents of all eukaryotic genomes. Satellite DNA contains repetitions of a basic monomeric unit, which is 180- to 360-bp long, and appears to be plentiful in heterochromatin in centromeric and telomeric regions [5]. Satellite DNAs are expanded on a chromosome by mutational processes such as replication slippage or unequal crossover which are likely to be the major driver of concerted evolution [5], and subsequently fixed. Interestingly, centromeric satellite DNA is known to have diverged significantly, not only between chromosomes within the same species, but also among a group of closely related higher eukaryotic species [5,6], suggesting rapid evolutionary patterns of centromeric DNA. Although no defined biological function has been established, satellite DNA may be involved in genome stability [5]. In particular, it has been demonstrated that centromeric satellite DNA is involved in centromere function via binding to centromere specific proteins, and the DNA and protein components specific to centromeric chromatin are evolving rapidly [5–7].

Another type of repetitive sequence, the mobile or transposable elements (TEs), is interspersed throughout the genome. TEs form a major component of plant genomes and are classified as class I, which transpose via an RNA intermediate, and class II, which transpose via a DNA intermediate. TEs contribute significantly to size variation between plant genomes. For the grasses, the portion of the genome contributed by long terminal repeat (LTR)-retrotransposons increases with genome size from rice, (14% of 430 Mbp) to maize (50–80% of 2500 Mbp) to barley (>70% of 4800 Mbp) [8]. TEs promote chromosomal rearrangements by transposition or illegitimate recombination as well as minor genomic variation through excision and insertion [9].

* Corresponding authors. Y.P. Lim is to be contacted at Department of Horticulture, College of Agriculture and Life Science, Chungnam National University, Daejeon 305-764, Republic of Korea. Fax: +82 42 823 1382. Y. Hur, Department of Biological Science, College of Biological Science and Biotechnology, Chungnam National University, Daejeon 305-764, Republic of Korea. Fax: +82 42 822 9690.
*E-mail addresses:* ykhur@cnu.ac.kr (Y. Hur), yplim@cnu.ac.kr (Y.P. Lim).
[1] These authors contributed equally to this paper.
[2] Present address.

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.12.002

*Brassica* species, within the Brassicaceae family, are economically important sources of vegetable oil, fresh and preserved vegetables, and condiments. The genetic relationship of the different diploid and amphidiploid cultivated *Brassica* species is described by the triangle of U [10]. Of the six widely cultivated species of *Brassica*, *B. rapa* (AA, 2n = 20), *B. nigra* (BB, 2n = 16), and *B. oleracea* (CC, 2n = 18) are monogenomic diploids. The remaining three species, *B. juncea* (AABB, 2n = 36), *B. napus* (AACC, 2n = 38), and *B. carinata* (BBCC, 2n = 34), are allopolyploids, which have evolved as a result of hybridization between different monogenomic diploids [10]. *Brassica* species are closely related to *A. thaliana* have diverged from a common ancestor 14.5–20.4 million years ago [11]. Comparative genetic and physical mapping between *Brassica* species and *A. thaliana* identified co-linear chromosome segments, conserved gene order, and a high degree of sequence conservation albeit with some variation in gene content through deletion, insertion, or dramatic chromosomal rearrangements and alterations since their divergence [12,13]. Compared to *A. thaliana*, however, the diploid *Brassica* genomes have been extensively triplicated with frequent genomic rearrangements [12,13] with an increased repetitive DNA fraction [12].

In addition to the comparative survey of coding regions between *Brassica* and *Arabidopsis* or among *Brassica* species, comparative analyses of the non-coding regions of the genome, represented primarily by repetitive DNA sequences, is of interest for the understanding of global genome evolution since *Brassica* species diverged from a common ancestor; however, little is known about the evolutionary relationship distribution and organization of repetitive elements within *Brassica* and closely related species. To address these questions, the isolation and characterization of corresponding sequences and assessment of their distribution within the related species are required. For global genome analysis, molecular cytogenetic methods, such as fluorescence *in situ* hybridization (FISH), are very powerful. Using FISH, the chromosomal localization of rDNAs [14], centromere-related tandem satellite DNAs (i.e. CentBr) [15–18], and B-genome-specific dispersed repetitive sequences [19], as well as the incorporated location of exogenous chromatin [20], have been studied in the *Brassica* genome. The Ty1/*copia*- and Ty3/*gypsy*-like retrotransposons have also been localized on chromosomes of both *B. oleracea* and *B. rapa* [17,21]. Moreover, this technology is likely to allow an insight into the evolution of repetitive DNAs among closely related species; changes in the number of rDNA loci in *Brassica* amphidiploids [14] and lack of CentBr repeats in *B. nigra* in contrast with their localization in pericentromeres of *B. rapa* and *B. oleracea* [18].

In this study, we report the molecular and cytogenetic characterization of the major repetitive DNAs that are conserved in pericentromeres, subtelomeres, and telomeres of *Brassica* species. Moreover, the conservation and divergence of the repetitive DNAs in the tribes Brassiceae (*Brassica* species and *Raphanus sativus*) and Camelineae (*A. thaliana*) were examined. Our results illustrate the organization and evolutionary dynamics of the primarily heterochromatic repetitive DNAs in *Brassica* species and the related species, *R. sativus*.

## 2. Materials and methods

### 2.1. Plant materials and genomic DNA isolation

Details of the *Brassica* species and *R. sativus* used in this study are listed in Table 1. Genomic DNA was isolated from young leaves using a standard CTAB (cetyltrimethylammonium bromide) protocol [22].

### 2.2. Isolation of repetitive DNA

Two *B. rapa* BAC clones (KBrH77C11 and KBrH77I01) that produced a ladder pattern that is typical of highly repetitive DNAs were identified during restriction digestion fingerprinting with *Hin*dIII (Supplementary data Fig. S1). These BAC clones also exhibited strong FISH signals at the pericentromeres of *B. rapa* (Fig. S2). In particular, KBrH77C11 and KBrH77I01 were found to hybridize specifically to eight and two chromosomes pairs, respectively, in a manner consistent with the chromosomal position pattern of CentBr1 and CentBr2 reported previously [17,18]. The two BAC DNAs were digested with *Hin*dIII and size-fractionated by electrophoresis to produce two fragments approximately 176 bp and 352 bp in size on an agarose gel. The BAC DNAs were also digested with *Bam*HI, *Eco*RI, and *Sau*3AI. After size-fractionation, the 176-bp and the 352-bp fragments were cloned. The tandem repeats cloned from KBrH77C11 and KBrH77I01 were designated as CentBr1 and CentBr2, respectively, according to a previous report [17].

Two *Ty1*/*copia*-like and *Ty3*/*gypsy*-like retrotransposons KBrH143L07e (GenBank accession number BZ613967) and KBrH143E08a (GenBank accession number BZ613814) were identified from a BAC end sequence database using BLAST [23]. The repeats were amplified by PCR using specific primers (KBrH143L07e: 5′-GCTTGG-CTGTGAACCTTGGA-3′ and 5′-GATCAAGTGTATACTTCCT-3′) and (KBrH143E08a: 5′-ATAGAGCTTGTGTTCCCGTA-3′ and 5′-TGCAC-CAGTCCTCTTTGTGA-3′) to produce 493-bp and 437-bp fragments, respectively. The products were cloned into the pGEM-T-Easy vector (Promega) according to the manufacturer's instructions, and these constructs were designated pBct and pBgt. A telomeric repeat oligonucleotide $(TTTAGGG)_n$ was synthesized according to [24].

A 348-bp *B. nigra* tandem repeat (GenBank accession number X16587) was identified from GenBank. This repeat was amplified by PCR, using the oligonucleotide primers 5′-GATCCTCCCCTTACATATTA-3′ and 5′-TTTTATTGCATAGTTGG-3′ from each of the diploid species studied. The amplified fragments from *B. rapa*, *B. nigra*, and *B. oleracea* were cloned into the pGEM-T-Easy vector (Promega) and designated p*Br*STR, p*Bn*STR, and p*Bo*STR, respectively.

### 2.3. Chromosome preparation and FISH

Details of the repetitive DNA, which was used for FISH probes, are listed in Table 2. Probes were labeled with digoxigenin-11-dUTP or biotin-16-dUTP by nick translation or PCR according to standard protocols (Roche). The preparation of mitotic metaphase chromosomes and FISH hybridization were as previously described [16].

### 2.4. Fiber-FISH

Leaf nuclei were prepared as described by Jackson et al. [25]. A suspension of nuclei was deposited at one end of a poly-L-lysine-coated slide (Sigma) and permitted to air-dry for 10 min. STE lysis buffer (8 μL) was pipetted onto the slide, and the sample was incubated at room temperature for 4 min. A clean cover-slip was used to slowly drag the

**Table 1**
Details of the *Brassica* species and *R. sativus* used in this study.

| Species | Genome type | Chromosome number (2n) | Accession number[a] |
|---|---|---|---|
| *B. rapa* ssp. *pekinensis* (Chinese cabbage) | AA | 20 | 241001 |
| *B. rapa* ssp. *oleifera* (turnip) | AA | 20 | 25023 |
| *B. rapa* var. brown sarson (brown sarson) | AA | 20 | 24026 |
| *B. nigra* (black mustard) | BB | 16 | 24018 |
| *B. oleracea* ssp. *capitata* (cabbage) | CC | 18 | 26036 |
| *B. oleracea* ssp. *italica* (broccoli) | CC | 18 | 26035 |
| *B. oleracea* ssp. *botrytis* (cauliflower) | CC | 18 | 26037 |
| *B. oleracea* ssp. *acephala* (kale) | CC | 18 | 26034 |
| *B. juncea* ssp. *varuna* (Indian mustard) | AABB | 36 | 24046 |
| *B. juncea* ssp. *rugosa* (Chinese mustard) | AABB | 36 | 26038 |
| *B. napus* (rapeseed) | AACC | 38 | 24034 |
| *B. carinata* (Abyssinian mustard) | BBCC | 34 | 24017 |
| *R. sativus* (radish) | RR | 18 | 26039 |

[a] The plant materials used in this study were donated by the Korea *Brassica* Genome Resource Bank (KBGRB) (http://www.brassica-resource.org/).

contents along the slide. The preparation was air-dried, fixed in ethanol: glacial acetic acid (3:1) for 2 min, and baked at 60 °C for 30 min. The slide was then incubated with a labeled probe mixture, covered with a 22-mm × 40-mm cover-slip and sealed with rubber cement. The slide was placed in an 80 °C oven in direct contact with a heated surface for 3 min and transferred to a wet chamber, pre-warmed in an 80 °C oven, for 2 min. The slide was then transferred to 37 °C for overnight incubation. Post-hybridization washing and signal detection were performed according to KOO et al. [16].

### 2.5. Southern hybridization

Genomic DNA was digested with *Hin*dIII or *Dra*I, fractionated on a 0.8% agarose gel, and transferred to a Hybond-N nylon membrane (Amersham Biosciences). The [α-$^{32}$P]-labeled probes were hybridized to the membranes in 5× SSC, 5× Denhardt's, with 0.5% SDS at 68 °C for 12 h. The membranes were washed twice with 2× SSC and 1% SDS at room temperature, followed by washing with 0.1× SSC and 0.1% SDS at 60 °C for 30 min. Signals were detected by autoradiography [22].

### 2.6. DNA sequencing and sequence analysis

Nucleotide sequences of CentBr, p*Br*STR, p*Bn*STR, and p*Bo*STR were determined using an automatic DNA sequencer (ABI377). Sequences were processed using phred [26] (i.e., base-calling, sequence quality check, and trimming of low-quality regions), and vector sequences were masked using cross-match (http://www.genome.washington.edu). After such pre-processing, redundant or partial (caused by sequence trimming) reads were removed by comparison with each other. Satellite repeat sequences were aligned using ClustalW (http://www.ebi.ac.uk/Tools/clustalw). A phylogenetic tree was generated using MEGA4 software (http://www.megasoftware.net) using the neighbor-joining method and the Tajima-Nei nucleotide substitution models. Bootstrap values were calculated from 1000 replicates. All satellite repeats sequenced have been submitted to GenBank (accession numbers EU294374–EU294389).

## 3. Results

### 3.1. Sequence divergence of 176-bp centromeric tandem satellite repeats (CentBr1 and CentBr2) found in B. rapa

*B. rapa* pericentromeres are known to be predominantly composed of two forms of 176-bp tandem repeats, termed CentBr1 (predominantly on eight chromosome pairs) and CentBr2 (on two chromosomes pairs) [17,18]. We performed a comparison of monomeric CentBr1 and CentBr2 sequences to investigate their divergence in the genome of *B. rapa*. We sequenced five 176-bp and one 352-bp fragments of *Hin*dIII-digested clones derived from KBrH077C11 and KBrH077I01, respectively. The 176 bp of monomeric repeats were compared with consensus CentBr1 and CentBr2 sequences generated from 198 490 *B. rapa* genome-wide BAC-end sequences, which were collected from the genome survey sequence (GSS) database of GenBank. The multiple sequence alignment revealed that the repeats derived from the distinct KBrH077C11 and KBrH077I01 were definitely classified into CentBr1 and CentBr2 clades, respectively (Fig. 1). Sequence identities in both clones were an average of 93.4% and 93.7%, respectively (Table S1); however, the similarity between the two clones was 84.6% (Table S1). This difference is due to the presence of SNPs in 10 positions, representing four transversions ($\alpha$) and six transitions ($\beta$) (Fig. 1). Moreover, CentBr2 repeats contained both *Hin*dIII (AAGCTT) and *Sau*3AI (GATC) endonuclease recognition sites, while the *Sau*3AI site was absent from CentBr1 (Fig. 1). In the restriction enzyme digestion assay, the BAC clone KBrH077I01 was digested by *Sau*3AI as well as *Hin*dIII, but KBrH077C11 was digested only by *Hin*dIII (Fig. S3), indicating a different biased distribution of CentBr repeats in the large inserted genomic DNAs. In conjunction with this finding, KBrH077C11 and KBrH077I01 were identified to specifically hybridize to eight and two chromosome pairs, respectively (Fig. S2). Thus, we suggest that the presence of the *Sau*3AI recognition site and the distinctive SNPs within the CentBr repeats can be an important hallmark for the divergence from the genome of *B. rapa*. Such differences between the CentBr1 and CentBr2 repeats were also identified in their dimeric repeats (352 bp; Fig. S4).

**Table 2**
Summary of repetitive DNAs used as probes in this study.

| Clone name | Source | Size (bp) | Remark[a] |
|---|---|---|---|
| CentBr1 | KBrH077C11[b] | 176 | *Brassica*/*Raphanus* centromeric satellite repeat (including *Hin*dIII site) |
| CentBr2 | KBrH077I01[b] | 176 | *Brassica*/*Raphanus* centromeric satellite repeat (including *Hin*dIII and *Sau*3AI sites) |
| pBct | KBrH143L07[b] | 493 | *Brassica*/*Raphanus* pericentromere-associated *Ty1*/*copia*-like retrotransposon |
| pBgt | KBrH143E08[b] | 437 | *Brassica*/*Raphanus* pericentromere-associated *Ty3*/*gypsy*-like retrotransposon |
| p*Br*STR | *B. rapa* genomic DNA | 352 | *B. rapa* subtelomeric satellite repeat |
| p*Bn*STR | *B. nigra* genomic DNA | 338–339 | *B. nigra* subtelomeric satellite repeat |
| p*Bo*STR | *B. oleracea* genomic DNA | 353–354 | *B. oleracea* subtelomeric satellite repeat |
| Telomere sequence | synthetic $(TTTAGGG)_n$ probe | | COX et al. (1993) |

[a] Described in *Results and discussions* of this article.
[b] Derived from the KBrH BAC library.

### 3.2. CentBr1 and CentBr2 are differentially distributed in Brassica species and R. sativus

To assess the broader conservation of CentBr tandem repeats, we examined various *Brassica* species and *R. sativus* using a combination of FISH and Southern blot hybridization (Table 3). The CentBr repeats are highly conserved in the closely related *B. rapa*/*B. oleracea* lineage, and this finding was confirmed with Southern blot hybridization data (Fig. S5). The CentBr repeats are also conserved in the genome of *R. sativus*, which belongs to the *B. rapa*/*B. oleracea* lineage [13], although the CentBr FISH signals were very weak (Fig. 2; Fig. S5). CentBr repeats were not identified in the genome of the more distantly related *B. nigra* or in the *B. nigra*-derived genomes of the amphidiploid species *B. juncea* and *B. carinata* (Fig. 2; Fig. S5). This result suggests that *Brassica* centromeric repeats have diverged during the evolution of the B and A/C-genome lineages.

An examination of CentBr repeats in different subspecies of *B. oleracea* revealed significant variation (Fig. 2; Table 3). CentBr FISH signals were observed on all chromosomes of cabbage and broccoli; however, these signals were absent from one chromosome pair of cauliflower and two chromosome pairs of kale (Figs. 2B–E). For CentBr1, cabbage, broccoli, and cauliflower exhibited strong FISH signals in seven of the nine chromosome pairs, while kale only exhibited strong signals in six chromosome pairs (Figs. 2B1–E1). CentBr2 FISH signals were also highly variable between *B. oleracea* subspecies, and the repeats were more abundant in cabbage and cauliflower (Figs. 2B2–E2). Different subspecies of *B. juncea* also showed a significant variation in CentBr FISH signals on the A-genome-derived chromosomes. CentBr1 signals in Chinese mustard were strongly detected on all 10 chromosome pairs derived from the A genome (Fig. 2G1); however, only 6 of the 10 A-genome-derived Indian mustard chromosome pairs demonstrated a strong signal, while the remaining four chromosome pairs showed weak hybridization (Fig. 2H1). CentBr2 FISH signals were also polymorphic between these species with two and one chromosome pairs hybridizing in Indian mustard (arrows in Fig. 2G2) and Chinese mustard, respectively (arrows in Fig. 2H2).

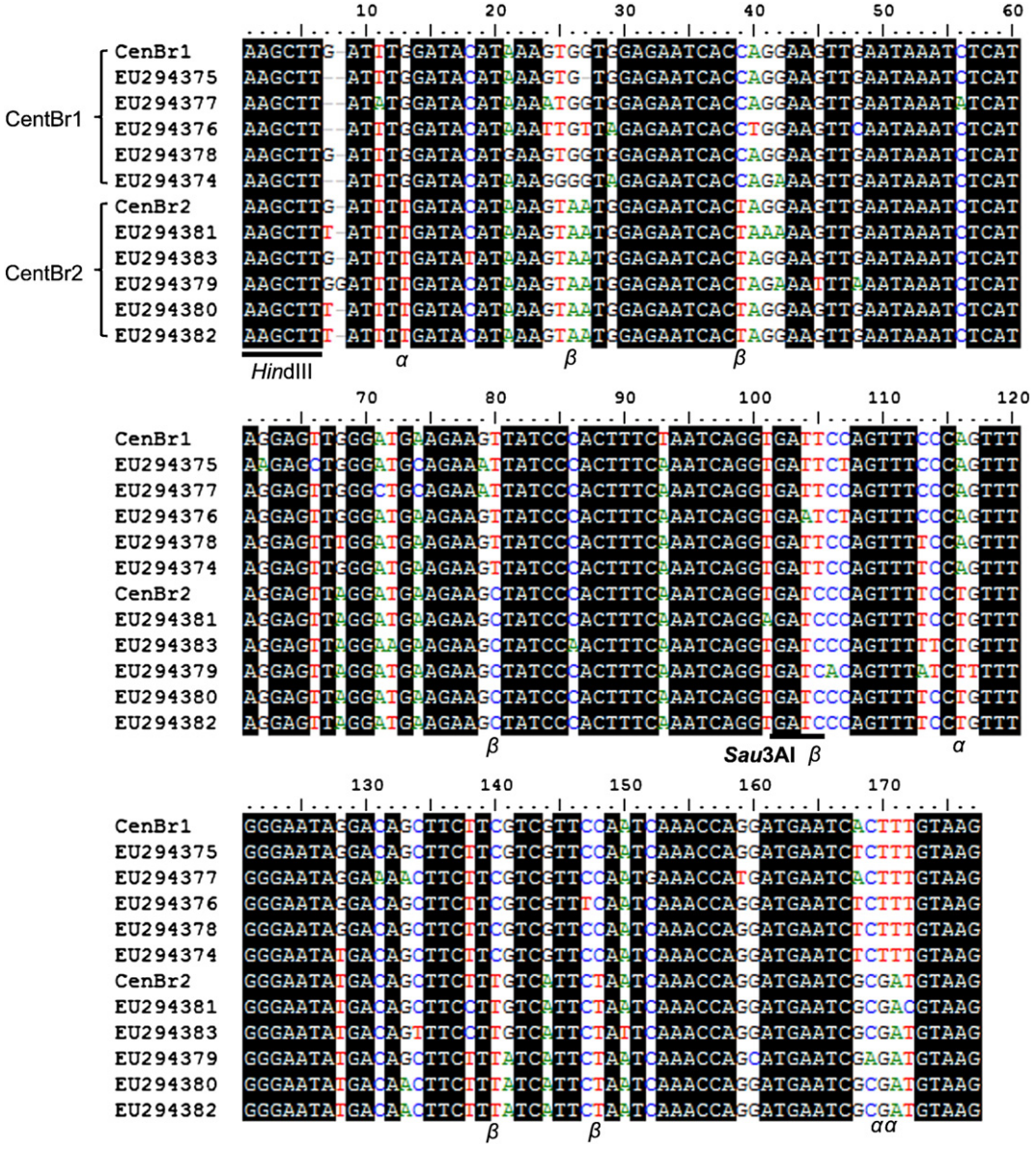


**Fig. 1.** Multiple-sequence alignment of CentBr1 and CentBr2 repeats. Conserved regions are shown in white on a black background. In the multiple alignment, nucleotide transversions and transitions are indicated by α and β, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 3**
The number of FISH hybridization sites in CentBr1, CentBr2, pBct, and pBgt identified on chromosomes of *Brassica* species and *R. sativus*.

| Species | CentBr1 | | | CentBr2 | | | pBct | | pBgt | |
|---|---|---|---|---|---|---|---|---|---|---|
| | S | W | N | S | W | N | S+W | N | S+W | N |
| *B. rapa* ssp. *pekinensis* | 16 | 4 | 0 | 4 | 16 | 0 | 20 | 0 | 6 | 14 |
| *B. rapa* ssp. *oleifera* | 16 | 4 | 0 | 4 | 16 | 0 | ND | 0 | ND | 0 |
| *B. rapa* var. brown sarson | 16 | 4 | 0 | 4 | 16 | 0 | ND | 0 | ND | 0 |
| *B. nigra* | 0 | 0 | 16 | 0 | 0 | 16 | 0 | 16 | 8 | 8 |
| *B. oleracea* ssp. *capitata* | 14 | 4 | 0 | 12 | 6 | 0 | 18 | 0 | 0 | 19 |
| *B. oleracea* ssp. *italica* | 14 | 4 | 0 | 16 | 2 | 0 | ND | 0 | ND | 0 |
| *B. oleracea* ssp. *botrytis* | 14 | 2 | 2 | 4 | 12 | 2 | ND | 0 | ND | 0 |
| *B. oleracea* ssp. *acephala* | 12 | 2 | 4 | 6 | 8 | 4 | ND | 0 | ND | 0 |
| *B. juncea* ssp. *varuna* | 12 | 8 | 16 | 4 | 16 | 16 | 32 | 4 | 2 | 34 |
| *B. juncea* ssp. *rugosa* | 20 | 0 | 16 | 2 | 18 | 16 | ND | 0 | ND | 0 |
| *B. napus* | 28 | 10 | 0 | 8 | 30 | 0 | 38 | 0 | 6 | 32 |
| *B. carinata* | 12 | 6 | 16 | 12 | 6 | 16 | 28 | 6 | 2 | 32 |
| *R. sativus* | 0 | 12 | 6 | 0 | 18 | 0 | 12 | 6 | 4 | 16 |

S, strong signal; W, weak signal; N, no signal; ND, not determined.

### 3.3. Sequence comparison of centromeric repeats found in different crucifer taxa

Centromeric satellite repeats of related species including *Brassica* species [27,28], *R. sativus* (177 bp) [29], *Diplotaxis erucoides* (176 bp) [30], *Sinapsis arvensis* (175 bp; from GenBank accession no. X74830), *A. thaliana* (178 bp of pAL1) [31], *A. arenosa* (179 bp of pAa) [31], *A. halleri* ssp. *gemmifera* (169 bp of pAge1 and 178 bp of pAge2) [32], *A. lyrata* (175 bp) [33], *Turritis glabra* (173 bp and 179 bp) [34], *Capsella rubella* (168 bp) [35], *Sisymbrium irio* (219 bp) [35], and *Olimarabidopsis pumila* (178 bp) [35], were identified within the GenBank nucleotide database and assembled to examine their divergence in the Brassicaceae family. The sequence comparison revealed that centromeric repeats in the tribe Brassiceae represented by *Brassica* spp., *D. erucoides*, *R. sativus*, and *S. arvensis*, are highly conserved with 71–85% sequence identity and similar repeat lengths. In contrast, centromeric repeats from *Arabidopsis* spp., *C. rubella*, *S. irio*, and *O. pumila* are highly divergent. Here, although *Sisymbrium* species were found to be clustered together with Brassiceae [36], no sequence similarity was found between *Brassica* species and *S. irio*, classifying them into different distant clades. These results were

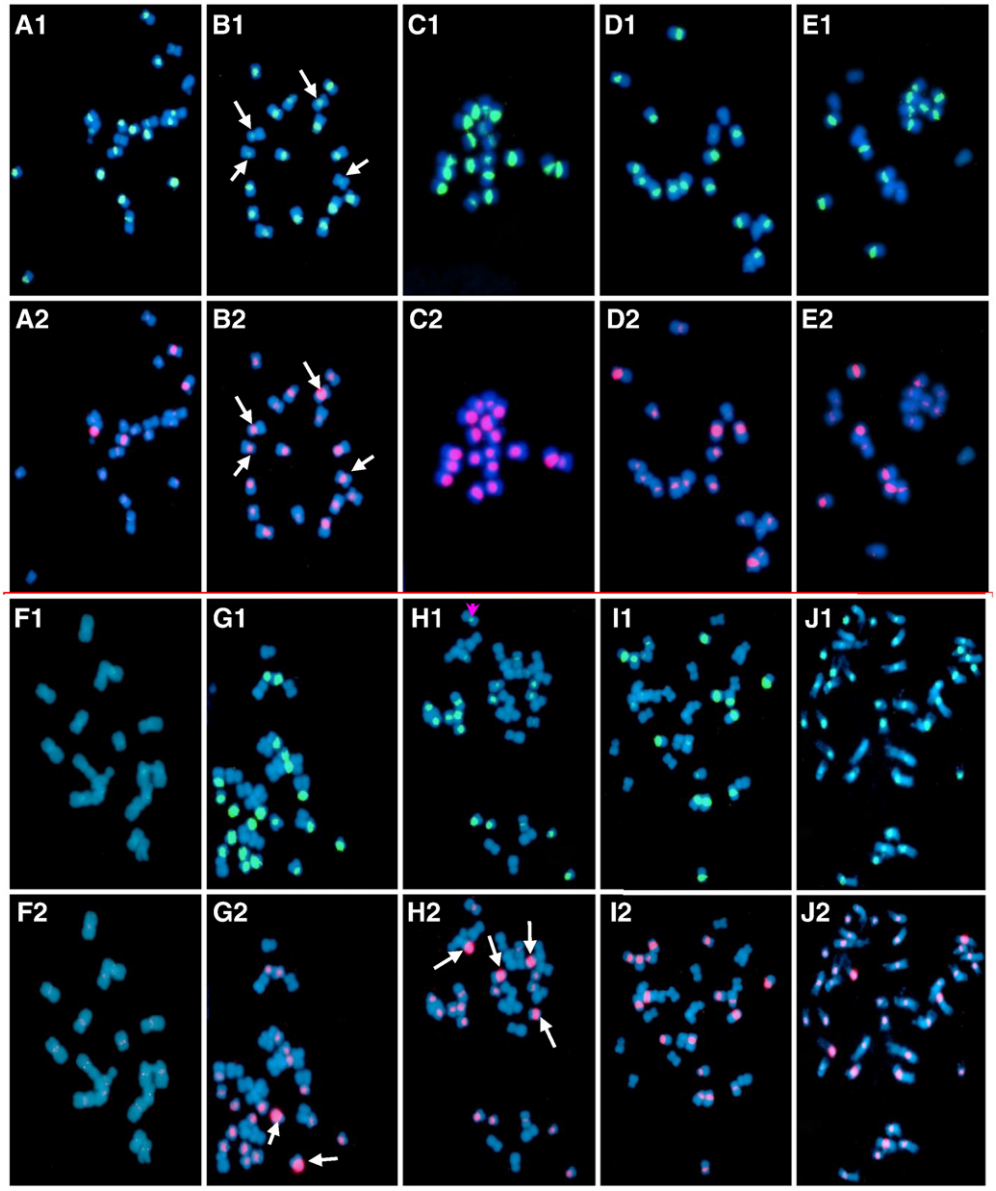


**Fig. 2.** FISH mapping of CentBr1 (green) and CentBr2 (red) repeats on mitotic metaphase chromosomes of *Brassica* species and *R. sativus*. (A) *B. rapa* ssp. *pekinensis*. (B) *B. oleracea* ssp. *capitata*. (C) *B. oleracea* ssp. *italic*. (D) *B. oleracea* ssp. *botrytis*. (E) *B. oleracea* ssp. *acephala*. (F) *R. sativus*. (G) *B. juncea* ssp. *rugosa*. (H) *B. juncea* ssp. *varuna*. (I) *B. carinata*. (J) *B. napus*.

demonstrated in the phylogenetic analysis, which resolved five distinct clades (Fig. 3).

Highly conserved sequence identity and similar repeat lengths (*Brassica* spp., *D. erucoides*, *R. sativus*, and *S. arvensis*), and highly divergent centromeric repeats (*Arabidopsis* spp., *C. rubella*, *S. irio*, and *O. pumila*) were identified by sequence comparison. This result demonstrates that there is a contrasting fashion between inter-tribal collinearity [13] and the dynamics of centromeric satellite DNAs. This result also showed that CentBr repeats rapidly evolved with conservation only within the tribe Brassiceae. Similar results have been reported previously for the grass family [7].

### *3.4. Chromosomal localization of rDNA sites in Brassica species and R. sativus*

The chromosomal localization of 5S and 45S rDNAs was identified in *Brassica* and *Raphanus* species via FISH mapping and the results are summarized in Table 4. The 5S and 45S rDNAs are localized in the pericentromeric heterochromatin regions of all the species, while 45S rDNAs are also localized in nucleolar organizer regions (NORs) (Table 4; Fig. S6). Each of the species demonstrated a distinct number of rDNA sites (Fig. S6). The distribution of 5S and 45S rDNA sites in subspecies of *B. oleracea* varied (Table 4). Six sites were identified in cabbage and cauliflower, of which, four sites were subtelomeric in the short arms of the acrocentric chromosomes and two sites were in the pericentromeric regions of the long arm of the NOR bearing chromosome (Table 4; Fig. S6). However, only the four subtelomeric sites were identified in broccoli and kale and the 45S rDNA sites were not detected in these subspecies.

The conservation of 5S or 45S rDNA sites was further studied in allotetraploid *Brassica* species. The relative number of 5S rDNA sites increased in *B. napus* across six chromosome pairs, while it decreased in *B. carinata* with loss of two 5S rDNA sites from the C genome (Table 4; Figs. S6F1 and S6G1). There was a general loss of 45S rDNA sites within *B. juncea* (Table 4; Fig. S6E2), while 45S rDNA sites in *B. napus* and *B. carinata* appear to be conserved or lost depending on the source of the

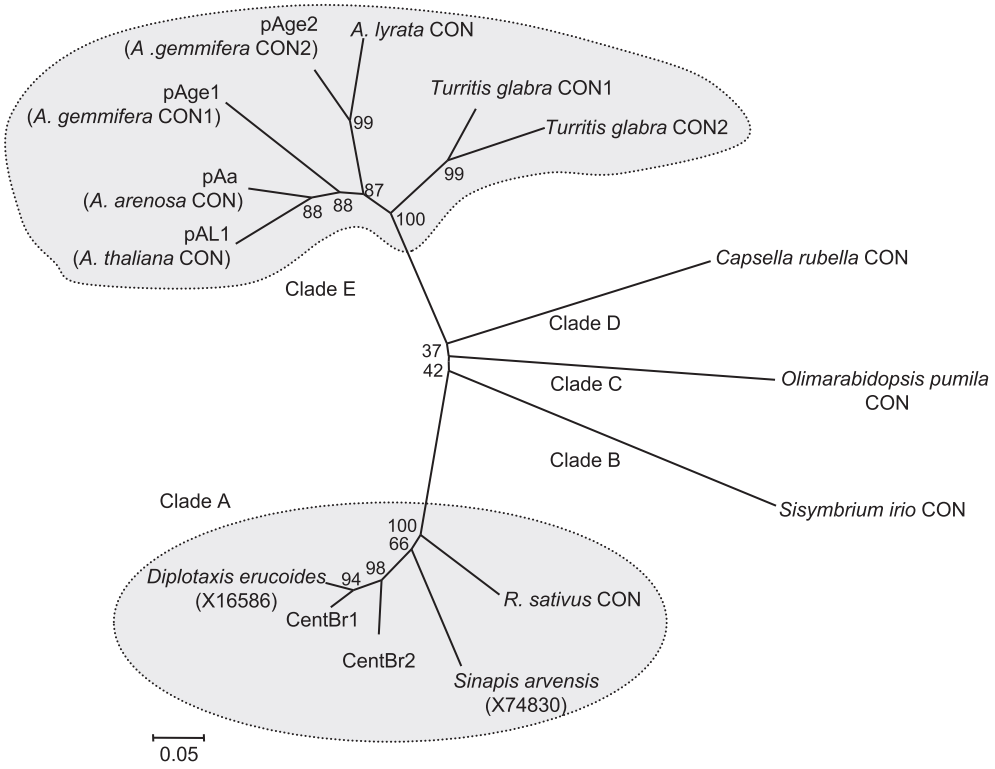


**Fig. 3.** Phylogeny of centromeric satellite repeats of nine genera belonging to Brassicaceae. After consensus, centromeric satellite repeats, which were determined by examining the most frequent nucleotide at each position of the centromeric satellite repeats of the corresponding species, were multiple-aligned. The phylogenetic tree was generated using the neighbor-joining method using the Tajima-Nei nucleotide substitution model.

diploid species (Table 4). The result shows that allotetraploid *Brassica* species appear to have undergone rapid genomic change, in rDNA sites, associated with allotetraploidization.

### 3.5. Identification of pericentromere-associated retrotransposons in Brassica and R. sativus

To examine the additional repeat structures within the Brassiceae genomes, we identified common retrotransposons from *B. rapa* BAC clones and studied their chromosomal locations using FISH. The BAC clone KBrH143L07 was observed to exclusively hybridize to pericentromeric regions of all 10 chromosome pairs, while KBrH143E08 hybridized to those of three chromosome pairs (Fig. S7). An end sequence (GenBank accession number BZ613967) derived from the BAC clone KBrH143E08 showed high sequence similarity (*E*-value of 5*e*-87 in TBLASTX search) with the CRB (Centromeric Retrotransposon of *Brassica*) polyprotein [18], a Ty1/*copia*-like retrotransposon. A 437-bp subsequence of this sequence was subcloned and designated pBct. Another end sequence (BZ613814) derived from the corresponding BAC clone also demonstrated high sequence similarity (*E*-value of 1*e*-101 in TBLASTX search) with the reverse transcriptase of PCRBr1b (Peri Centromeric Retrotransposons of *B. rapa*-1b), a Ty3/*gypsy*-like retrotransposon [18]. A 471-bp subsequence was subcloned and designated as pBgt.

Conservation and chromosomal location of these two retrotransposons were determined in *Brassica* species and *R. sativus* using Southern blot hybridization and FISH (Table 3). These studies showed that pBct is highly conserved in the centromere-proximal regions of *Brassica* A/C genome species and *R. sativus* (Fig. 4; Fig. S5);

**Table 4**
Location and number of 5S and 45S rDNA sites for *Brassica* species and *R. sativus*.

| Species | No. of chromosomes with 5S rDNA (No. of sites) | Location of 5S rDNA (No. of sites) | No. of chromosomes with 45S rDNA sites (No. of sites) | Location of 45S rDNA (No. of sites) | No. of chromosomes with either two 5S or two 45S rDNA loci closely adjacent in a chromosome | No. of chromosomes with 5S and 45S rDNA loci in the same chromosome |
|---|---|---|---|---|---|---|
| *B. rapa* ssp. *pekinensis* | 6 (6) | PH[a] (6) | 10 (10) | NOR[b] (2) + PH (8) | – | 4 |
| *B. nigra* | 2 (2) | PH (2) | 6 (6) | NOR (4) + PH (2) | – | – |
| *B. oleracea* ssp. *capitata* | 2 (4) | PH (4) | 4 (6) | NOR (4) + PH (2) | 2 (45S rDNAs), 2 (5S rDNAs) | – |
| *B. oleracea* ssp. *botrytis*[c] | 2 (4) | PH (4) | 4 (6) | NOR (4) + PH (2) | 2 (45S rDNAs), 2 (5S rDNAs) | – |
| *B. oleracea* ssp. *italica*[c] | 2 (4) | PH (4) | 4 (4) | NOR (4) + PH (0) | 2 (5S rDNAs) | – |
| *B. oleracea* ssp. *acephala*[c] | 2 (4) | PH (4) | 4 (4) | NOR (4) + PH (0) | 2 (5S rDNAs) | – |
| *B. juncea* ssp. *varuna* | 8 (8) | PH (8) | 14 (14) | NOR (6) + PH (8) | – | 4 |
| *B. juncea* ssp. *rugosa*[c] | 8 (8) | PH (8) | 14 (14) | NOR (6) + PH (8) | – | – |
| *B. napus* | 12 (16) | PH (16) | 14 (14) | NOR (6) + PH (8) | 4 (5S rDNAs) | 12 |
| *B. carinata* | 4 (4) | PH (4) | 10 (10) | NOR (8) + PH (2) | – | – |
| *R. sativus* | 4 (4) | PH (4) | 6 (6) | NOR (2) + PH (4) | – | 2 |

[a] PH: pericentromeric heterochromatin.
[b] NOR: nucleolar organizer region.
[c] Not shown in Supplementary Fig. S6.

however, while pBct hybridization was detected in *B. nigra* using Southern hybridization, no hybridization was detected using FISH, reflecting the lower sensitivity of the FISH method and the low abundance of pBct in the *B. nigra* genome (Fig. S5). In contrast, pBct hybridization was detected in the centromere-proximal regions of the B genome chromosomes in *B. juncea* and *B. carinata* (Figs. 4G and I). The pBgt was highly conserved in centromere-distal regions of *Brassica* species and *R. sativus* with the exception of the C-genome (Fig. 4; Fig. S5). Interestingly, the number of pBgt sites observed in *B. juncea* and *B. carinata* was reduced compared with the number of sites in the two diploids, *B. rapa* and *B. nigra* (Figs. 4J and L; Table 3). This result suggests that genomic changes are associated with polyploidization.

The location of pBct and CentBr repeats were also analyzed using fiber-FISH (Fig. 4M), and these studies revealed that pBcts are inserted into CentBr arrays. The size of the pBct repeats was estimated to be approximately 3–6 kb (representing ≈1–2 μm in fiber-FISH; Fig. 4M). Co-signals for pBgt and CentBr were either very weak or not observed.

### 3.6. Identification of subtelomeric tandem repeats in diploid Brassica species

We recently identified an uncharacterized tandem repeat sequence in *B. nigra* (GenBank accession number X16587; 348 bp) within the GenBank nucleotide database by sequence identity with known subtelomeric repeats. PCR analysis demonstrated that this repeat sequence is conserved across the *Brassica* A, B, and C genomes; however,

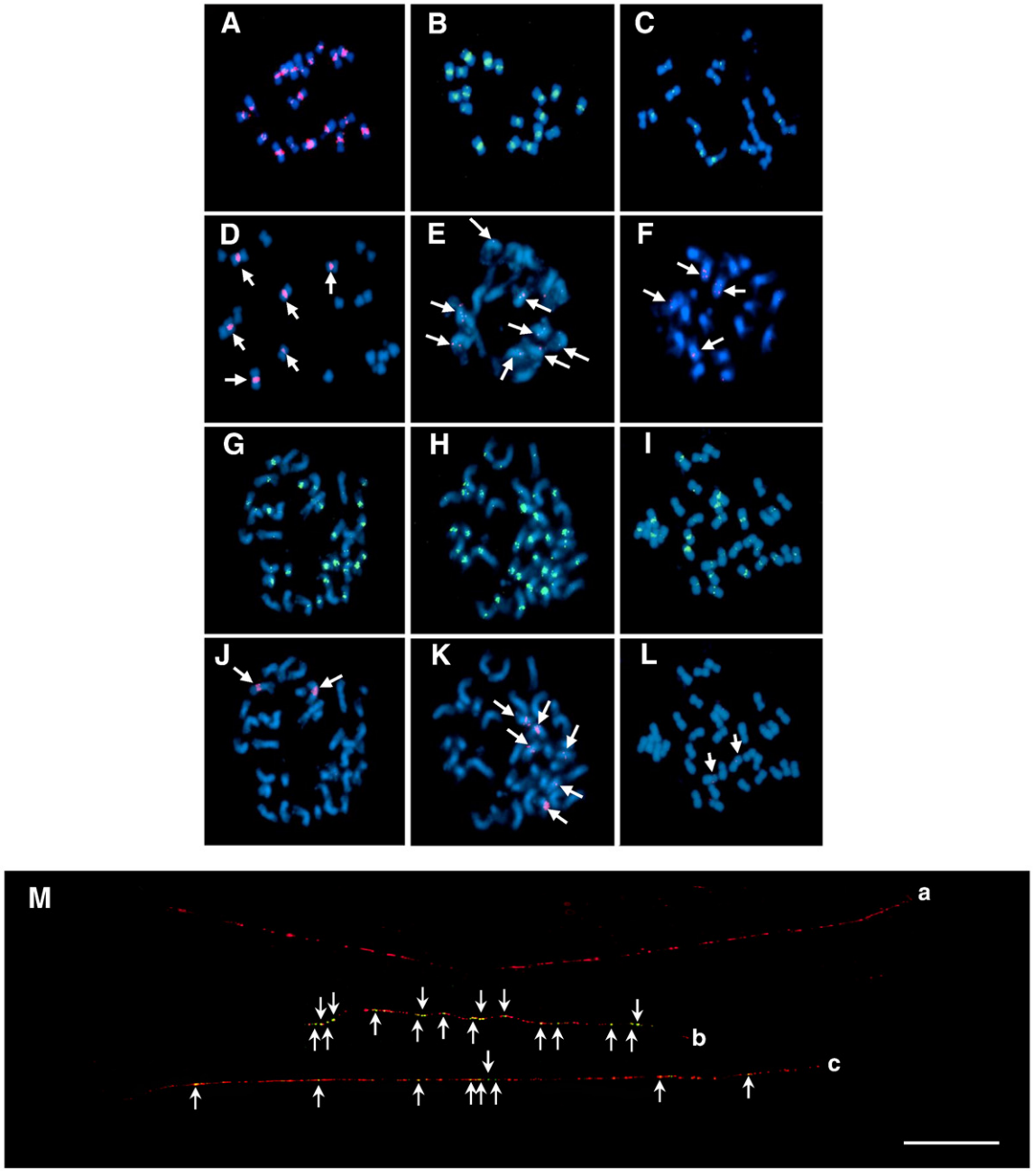


**Fig. 4.** FISH mapping of pBct and pBgt on the mitotic metaphase chromosomes of *Brassica* species and *R. sativus*. (A) and (D) *B. rapa* ssp. *pekinensis*. (B) *B. oleracea* ssp. *capitata*. (C) and (F) *R. sativus*. (E) *B. nigra*. (G) and (J) *B. juncea* ssp. *varuna*. (H) and (K) *B. napus*. (I) and (L) *B. carinata*. The pBct repeats were hybridized to (A)–(C) and (G)–(I) chromosomes, and pBgt repeats were hybridized to (D)–(F) and (J)–(L) chromosomes. (M) Fiber-FISH analysis conducted by hybridizing the probes CentBr1 (red) and pBct (green) to extended DNA fibers prepared from *B. rapa*. Fiber-FISH signals indicate highly variable densities of pBct within the CentBr1 arrays. The yellow arrows indicate FISH signals for pBct or pBgt.

the repeat failed to amplify from *R. sativus* or *A. thaliana*, yielding only faint and smeary products (Fig. S8). The PCR amplification products were cloned, sequenced, and designated as p*Br*STR, p*Bn*STR, and p*Bo*STR. Sequence similarity among the three repeats was relatively high (approximately 76–87%), with 27% GC content (Fig. 5A; Table S2). Among these repeats, the regions from 1 to 60 bp and 321 bp to the end of the region were highly conserved. The intermediate region, especially from 80 bp to 320 bp, however, contained significant nucleotide substitutions and a few insertions/deletions (InDels; Fig. 5A). These nucleotide changes may be responsible for the divergence of p*Bn*STR and other repeats (Fig. 5B). Each repeat was hybridized to genomic DNA from each of the four species, *B. rapa, B. nigra, B. oleracea,* and *R. sativus,* and then digested with *Dra*I. The resultant ladder pattern resembles the typical tandem repeat DNA structure (Fig. 6A). Under stringent

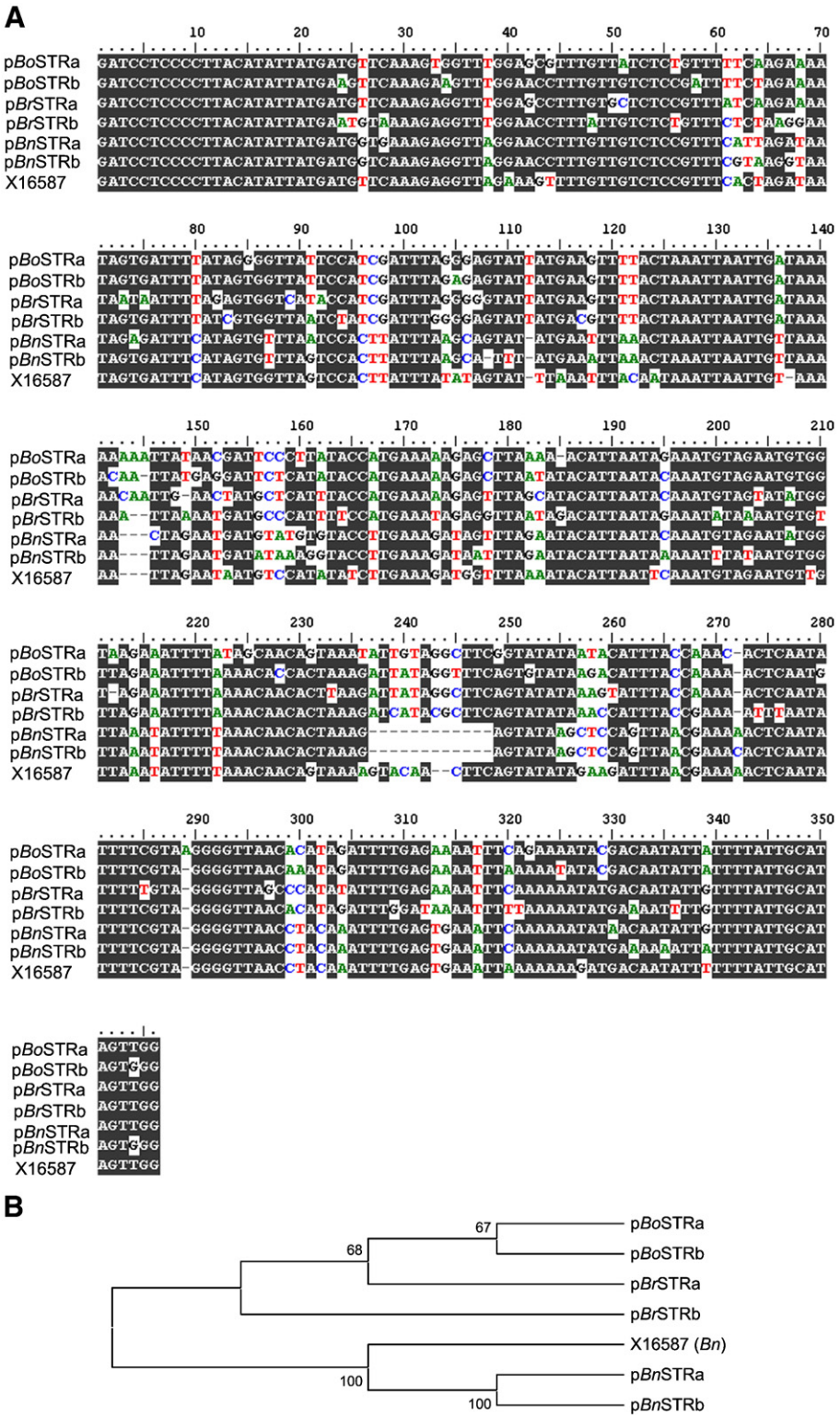


**Fig. 5.** Multiple-sequence alignment (A) and phylogeny (B) of tandem satellite repeats, p*Br*STR, p*Bn*STR, and p*Bo*STR, isolated from *B. rapa* ssp. *pekinensis*, *B. nigra*, and *B. oleracea* ssp. *capitata*, respectively.

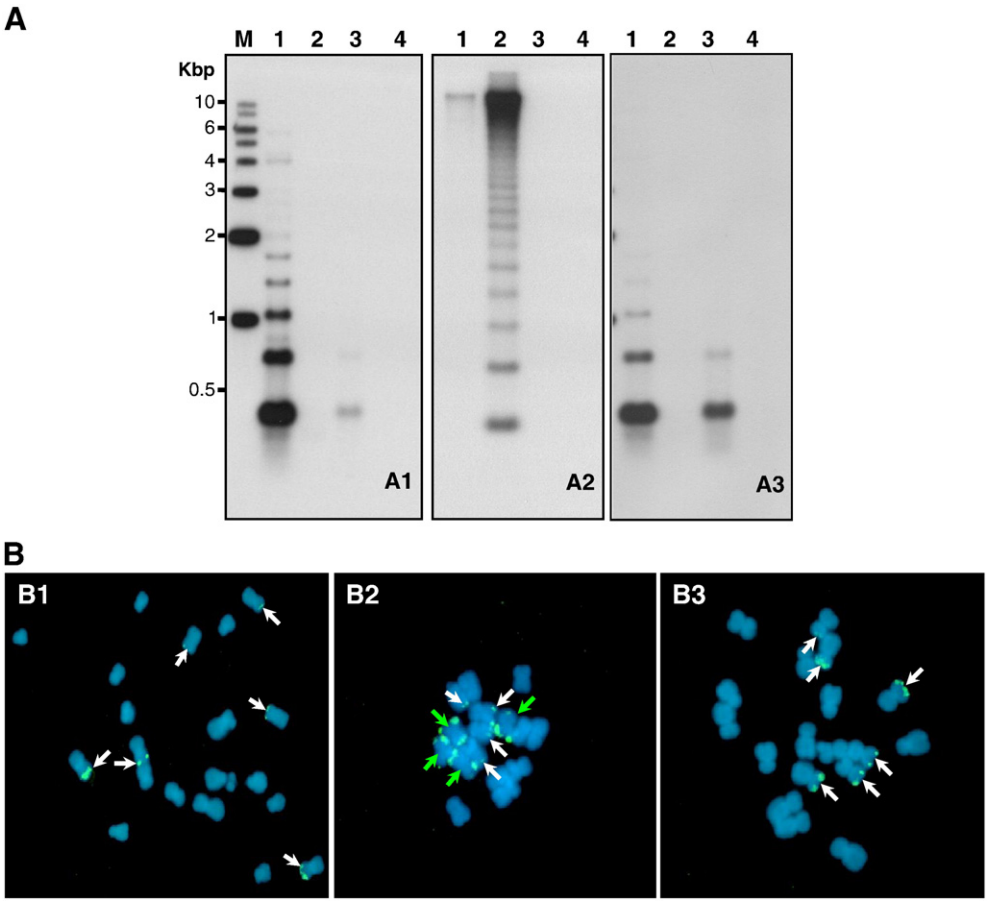


**Fig. 6.** Genomic organization and chromosomal localization of p*Br*STR, p*Bn*STR, and p*Bo*STR. (A) The following samples were used for Southern blot hybridization: (M) size marker, (1) *B. rapa* ssp. *pekinensis*, (2) *B. nigra*, (3) *B. oleracea* ssp. *capitata*, and (4) *R. sativus*. Genomic DNAs were digested with *Dra*I, separated, and transferred onto nylon membranes. The filters were hybridized with p*Br*STR (A1), p*Bn*STR (A2), and p*Bo*STR (A3). (B) Chromosomal localization of p*Br*STR, p*Bn*STR, and p*Bo*STR were performed on metaphase chromosomes of *B. rapa* (B1), *B. nigra* (B2), and *B. oleracea* (B3), respectively. The white arrows indicate FISH signal on one arm of a chromosome, and the green arrows indicate FISH signal on both arms of a chromosome.

hybridization conditions, p*Br*STR was hybridized to *B. rapa* (Fig. 6A1), while p*Bn*STR specifically hybridized to *B. nigra* (Fig. 6A2). No hybridization to *R. sativus* was detected; however, p*Bo*STR was simultaneously specific for both *B. rapa* and *B. oleracea*. Each repeat was also hybridized to mitotic metaphase chromosomes of the corresponding species and detected using FISH. Both p*Br*STR and p*Bo*STR were localized to the subtelomeric regions within one arm of three chromosome pairs for *B. rapa* and *B. oleracea*, respectively (Figs. 6B1 and B3). *B. nigra* p*Bn*STR hybridized to subtelomeric regions on one arm of two chromosome pairs as well as within both the long and short arms for two chromosome pairs (Fig. 6B2). Neither p*Br*STR nor p*Bo*STR hybridized to chromosomes of *B. nigra*.

The divergence of the p*Br*STR repeats in the *B. rapa* genome was examined by analyzing 198, 490 BAC-end sequences representing a random sampling of the genome. The p*Br*STR repeats matched 826 (0.42%) of the total BAC-end sequences. Of the identified sequences containing the repeat, 350 bp of 420 repeats were selected, assembled using ClustalW, and further analyzed with the MEGA software. The sequence comparison revealed high sequence identity (an average of 89% [SD 5%]) and a close nucleotide pairwise distance (0.094). Despite this high sequence conservation, nucleotide substitutions and InDels were observed at many different positions in the repeats. Phylogenetic analysis of the repeats identified two major clades representing class I and class II repeats (Fig. 7). The class I repeats most commonly represented 384 of the 420 repeats studied (91.4%), whereas the class II repeats were not commonly observed, representing only 36 of the 420 repeats (8.6%). The class I repeats appear to have extensively diverged.

### 3.7. Characterization of telomeric repeats in diploid Brassica species and R. sativus

Telomeres of diploid *Brassica* species and *R. sativus* were analyzed by Southern blotting with the *Arabidopsis*-type telomere repeat sequence $(TTTAGGG)_n$. Telomere size was found to vary between species. The ranges of telomere length were estimated to be 1.5 to 5.0 kb for *B. rapa*, 2.0 to 3.0 kb for *B. nigra*, 1.2 to 6.0 kb for *B. oleracea*, and 1.5 to 3.0 kb for *R. sativus* (Fig. 8A). The length of the *A. thaliana* telomere was estimated to range from 2.0 to 5.0 kb, a value that is consistent with previous data [37] (Fig. 8A). This result demonstrates that telomere size has changed through evolution within *Brassica* species.

Telomeric regions within the diploid *Brassica* species and *R. sativus* were also characterized using FISH (Fig. 8B). FISH signals were relatively weak in *B. rapa, B. nigra,* and *R. sativus* but were moderate in *B. oleracea*, and these results correspond to the greater telomere lengths in *B. oleracea*. Interestingly, strong telomeric repeat signals were also identified in the interstitial regions of six chromosomes (arrows in Fig. 8B5) of *B. oleracea*. Of these six chromosomes, three exhibited interstitial telomere-like repeats (ITRs) on both chromosome arms (arrows in Fig. 8B5). In connection with localization of telomeric repeats in interstitial regions, a homology-based search also revealed that ITRs were co-distributed with CentBr repeats in large insert DNAs of *B. oleracea* (GenBank accession number: AC183496) and *B. napus* (AC236792) (Fig. S9), respectively. This suggests a localization of ITRs in pericentromeres of *Brassica* species. This result is in contrast to Hasterok et al. [20], who found the *Arabidopsis*-type telomeric repeat hybridized

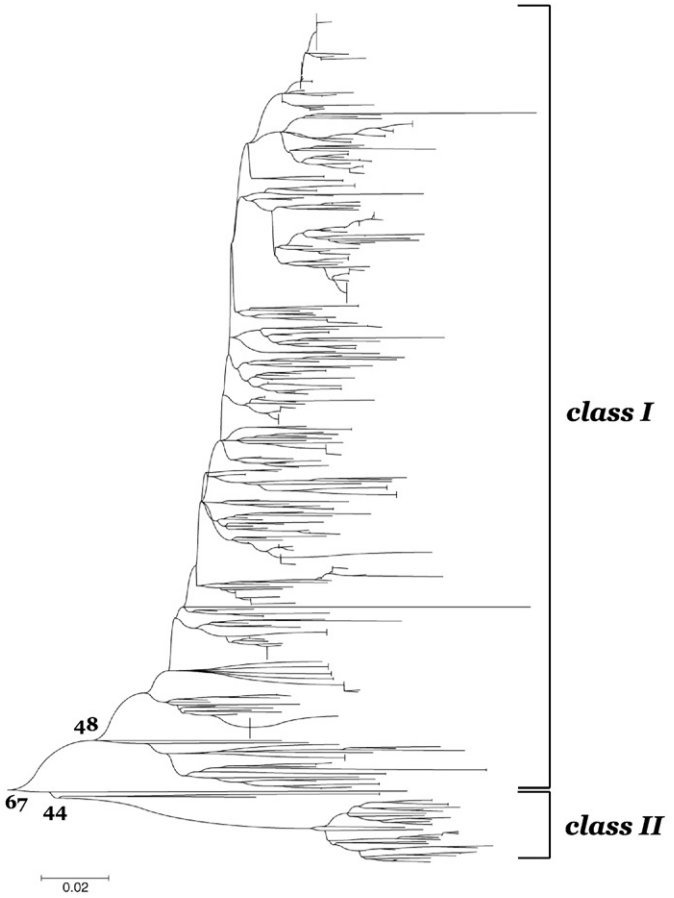


**Fig. 7.** Divergence of p*Br*STR repeats in the *B. rapa* genome revealed by phylogenetic analysis. The phylogenetic tree was generated using 420 copies of 350-bp repeats by the neighbor-joining method with the Kimura two-parameter option. Boot-strap values were calculated from 1000 replicates.

exclusively to the chromosome termini. This finding may be due to the greater sensitivity of repeat detection in our experiment.

## 4. Discussion

Repetitive DNA sequences, including tandem repeats, dispersed repeats, retroelements, and telomeric repeats, are responsible for the heterogeneity in genome size and composition. Thus, repetitive DNA is a major contributor to plant chromosome structure. The repetitive DNA in the genome is also important for evolutionary, genetic, taxonomic, and applied studies. Evolutionary relationships among plant species can be investigated by examination of the repetitive sequences. Thus, we examined the molecular and cytogenetic characteristics of the major repetitive DNA sequences on pericentromeres, subtelomeres, and telomeres, in order to study the relationship among *Brassica* relatives, *R. sativus*, and *A. thaliana*.

### *4.1. Centromeric tandem satellite repeats*

Centromeres in *B. rapa* are represented by CentBr1 and CentBr2 [17,18]. The repeats have diverged as evidenced by the loss of the intrinsic specific enzyme recognition site (*Sau*3AI) in CentBr1 and some nucleotide substitutions. In addition, their biased distribution on *B. rapa* chromosomes has been identified. In this study, we demonstrated significant variation of CentBr repeats in different *Brassica* species, suggesting that the *Brassica* centromeres may have undergone relatively rapid evolution within the different genomes. The observation of the intensity of FISH signals in the major *Brassica* species revealed that CentBr1 repeats were relatively more abundant than CentBr2 repeats (Fig. 2) except for in the *B. oleracea* ssp. *italica*, *B. carinata,* and *R. sativus*, indicating that the former repeat is the most common type in the majority of *Brassica* species. Interestingly, no CentBr repeats were detected in the *Brassica* B-genome, probably implying the accumulation of extensively diverged centromeric satellite repeats in the genome. We also found that while CentBr hybridization pattern was uniform across different subspecies of *B. rapa* (*B. rapa* ssp. *pekinensis*, *B. rapa* ssp. *oleifera,* and *B. rapa* var. brown sarson), significant variation of CentBr FISH signals were detected in different four subspecies of *B. oleracea* analyzed (Table 3). In general, CentBr1 repeats appear to be highly abundant in the four subspecies as shown in *B. rapa*. However, in comparison with FISH patterns in *B. oleracea* ssp. *capitata* and *B. oleracea* ssp. *italica*, the number of CentBr1 repeat site was reduced in *B. oleracea* ssp. *botrytis* (a pair of chromosome) and *oleracea* ssp. *acephala* (two pairs of chromosomes), indicative of the quantitative reduction of CentBr1 repeats in the two subspecies. Interestingly, the number of CentBr2 repeat site was highly variable among the four subspecies. In particular, *B. oleracea* ssp. *capitata* and *B. oleracea* ssp. *italica* have more than two times as many sites exhibiting strong FISH signals as *B. oleracea* ssp. *botrytis* and *oleracea* ssp. *acephala*. In particular, CentBr2 repeats in *B. oleracea* ssp. *italica* are likely to be a few more abundant than CentBr1 repeats in it. Moreover, there is a decreasing preference for chromosomal distribution of CentBr1 and CentBr2 repeats. Such expansion of CentBr2 repeats is likely to be species-specific. Additionally, variation of CentBr2 repeat sites has nothing to do with known phylogenetic relationship among the four subspecies. Besides diploid species, of the amphidiploids, different subspecies of *B. juncea* showed variation of CentBr repeat sites on A-genome-derived chromosomes. These results indicate that although CentBr repeats are highly conserved in *Brassica* and *Raphanus* species, their abundances in centromere regions are very polymorphic among species or chromosomes. This result implies that the divergence of CentBr may result from concerted evolution of repeats and their fixation in a specific genome or chromosomes [38]. The presence of polymorphic subclasses of centromeric repeats has also been reported in *Arabidopsis*, maize, and rice where the centromere repeats were found to be differentially localized to different chromosomes [35,39,40].

### *4.2. Pericentromere-associated retrotransposons*

CRB and PCRBr retrotransposons have previously been identified in the near centromeric regions of *Brassica* genomes [18]. Here, we demonstrated that pBct and pBgt have diverged from the previously characterized CRB and PCRBr, respectively, and have a distinct distribution pattern, despite the high sequence similarity. When pBct and CRB were compared, CRB was found in all centromeres of the genus *Brassica* [18], while pBct was not detected on some of the B-genome-related chromosomes within *B. juncea* and *B. carinata* (Figs. 4G and I; Table 3). This result indicates that pBct is likely to have been lost from some chromosomes of the *Brassica* B-genome following the divergence of the *Brassica* A/C and B genome lineages. PCRBr was previously detected in pericentromeric heterochromatin blocks of four chromosome pairs of *B. rapa*, *B. juncea*, and *B. napus*, indicating that this sequence was specific to the *Brassica* A-genome [18]; however, we found that pBgt was conserved in the pericentromeric regions of *B. rapa, B. nigra,* and *R. sativus* but not in *B. oleracea* (Fig. 4; Fig. S5). From these data, we suggest that pericentromeric retrotransposons such as CRB, pBct, PCRBr, and pBgt have been differentially accumulated in the genomic regions of different species after their divergence from the common ancestor.

We also demonstrated that pBcts are inserted into the CentBr repeat arrays of *B. rapa*. This result is consistent with previous studies of *Arabidopsis*, rice, and maize [39–41], suggesting that insertion events of pericentromere-specific retrotransposons into centromeric tandem repeat arrays may contribute to substantial structural changes in the centromeric regions during evolution. Unlike pBcts, pBgts were rarely inserted into dense CentBr repeat arrays, indicating their presence in centromere-distal regions. This investigation will offer insight into

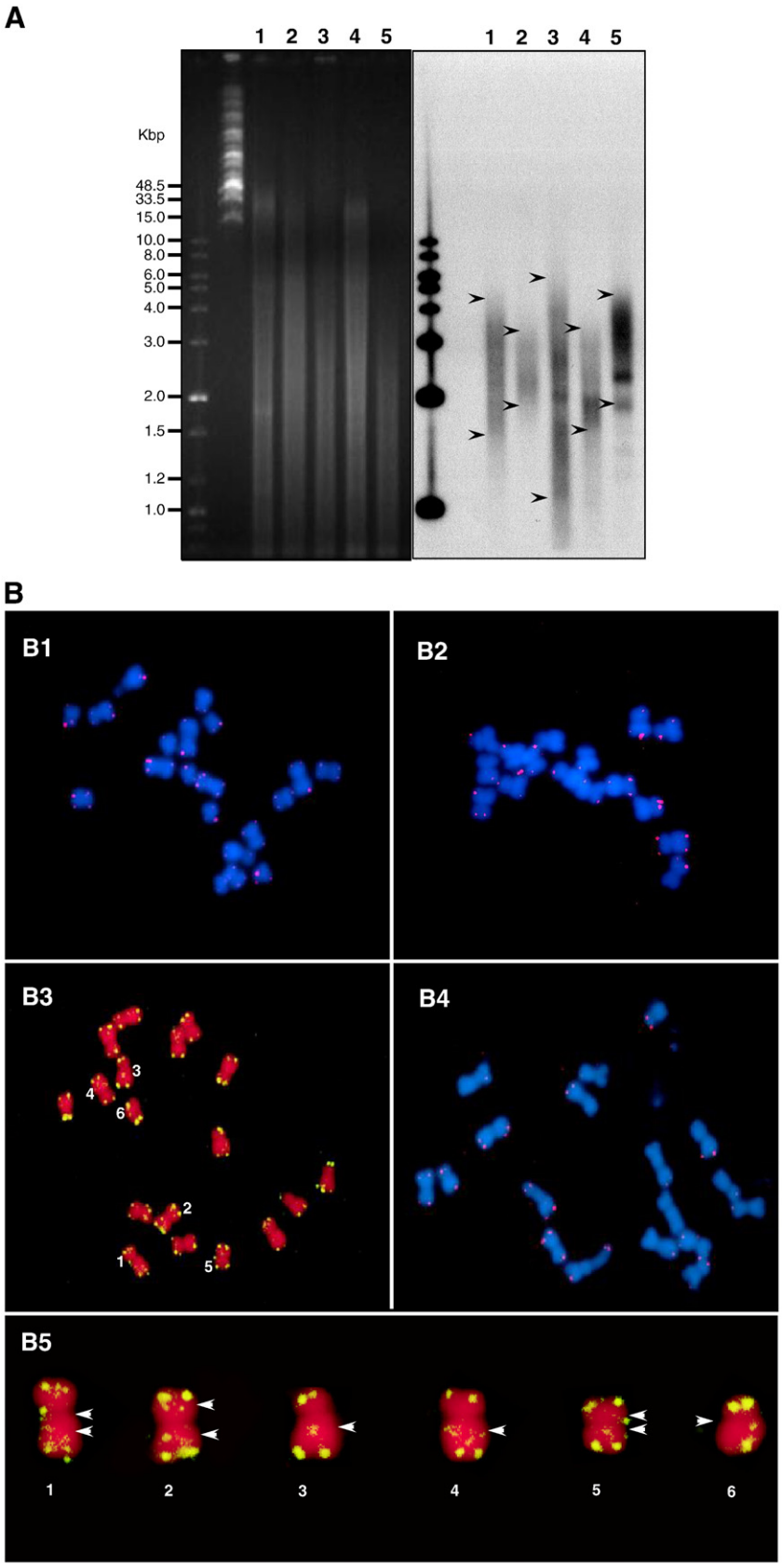


**Fig. 8.** Identification of length (A) and chromosomal location (B) of telomeres in diploid *Brassica* species and *R. sativus*. (A) The following samples were subjected to Southern blot hybridization: (1) *B. rapa* ssp. *pekinensis*, (2) *B. nigra*, (3) *B. oleracea* ssp. *capitata*, (4) *R. sativus*, and (5) *A. thaliana*. Each genomic DNA sample was digested with *Taq*I, separated, and transferred to nylon membranes. Filters were hybridized with the synthetic telomere probe $(TTTAGGG)_n$. (B) For FISH, the synthetic telomere probe was hybridized to the metaphase chromosomes of *B. rapa* ssp. *pekinensis* (B1), *B. nigra* (B2), *B. oleracea* ssp. *capitata* (B3), and *R. sativus* (B4). (B5) The arrowheads indicate interstitial telomere-like repeats (ITRs) observed on the metaphase chromosomes of *B. oleracea*.

understanding the structural divergence of centromeres of *Brassica* species and inferring their evolution.

### 4.3. Subtelomeric tandem repeats

In this study, we identified the novel subtelomere-associated satellite repeats p*Br*STR, p*Bn*STR, and p*Bo*STR in the diploid *Brassica* species. The repeats were not conserved in the close relative *R. sativus* or in *A. thaliana*, indicating that these repeats are species-specific tandem repeat DNA. Interestingly, although relatively high sequence similarity was found among the repeats, these sequences appear to have evolved in a genome-specific manner with characteristic sequence variation among *Brassica* A, B, and C genomes. The p*Bo*STR repeat was conserved only in the C and A genomes, while the p*Bn*STR repeat was found only in the B genome. This result implies that these two repeats have diverged according to the *Brassica* lineage. In comparison with p*Bo*STR, p*Br*STR was present only in the A genome, indicating that this sequence is another type of diverged class and an A-genome-specific repeat. This notion was supported by phylogenetic analysis for p*Br*STR repeats, which revealed that these repeats have extensively diverged in the A-genome with two major clades. Interestingly, our FISH analysis for p*Br*STR and p*Bo*STR repeats revealed that these sequences co-localized on the subtelomeric regions within one arm of three chromosome pairs for *B. rapa* and *B. oleracea*, suggesting that p*Br*STR and p*Bo*STR repeats are major repeats of the subtelomeres of the A-genome. Based on these results, we suggest that the diversity of these repeats shows independent evolution in the specific lineage and genome and concerted evolution of repetitive DNAs. In accord with these results, similar studies have described lineage- or genome-specific divergence of subtelomeric repeats in *Oryza* (TrsA repeat) [42] and *Aegilops* species (pGc1R-1 repeat) [43].

We also identified the type and size of telomeric repeats in *Brassica* and *Raphanus* species. The telomeric DNAs of *Brassica* and *Raphanus* species were composed of *Arabidopsis*-type TTTAGGG-type repeats. The *Arabidopsis*-type telomere has been found in most flowering plants [1,44]; however, several reports indicated that the repeat motif was not ubiquitous. For example, in the Alliaceae family, an alternate to the TTAGGG motif was found, and in other cases, loss of the TTAGGG telomeric sequence was identified [45,46]. Although the *Arabidopsis*-type repeat is well-conserved in *Brassica* and *Raphanus* species, telomere sizes varied between those species, and this size variation had no correlation with the genome sizes. Interestingly, in comparison with the telomere sizes of cereals (12–15 kb), tomato (30–60 kb), and tobacco (60–160 kb) [1,44], Brassiceae and *Arabidopsis* appear to have relatively short telomeres with approximately 1.2 to 6.0 kb of DNA. Additionally, comparison of FISH signals in the telomeres of *Brassica* and *Raphanus* species showed that a predominant abundance of telomeric sequences was found in *B. oleracea*.

Interestingly, unlike *B. rapa*, *B. nigra*, and *R. sativus*, the telomeric sequences of *B. oleracea* were observed in some interchromosomal regions (Fig. 8C1). The presence of interstitial telomere-like repeats (ITRs) in interchromosomal regions of *B. oleracea* indicated that the *B. oleracea* genome has undergone chromosomal rearrangements such as end-fusions or segmental duplications [47], which would be a very effective process in chromosome evolution leading to speciation. In addition, Biessmann et al. [48] proposed that ITRs are correlated with general genomic instability, including the creation of recombination hotspots, chromosomal breakage, and subsequent telomere-mediated healing. Therefore, the observation of ITRs in *B. oleracea* provides important evolutionary evidence for past telomere-mediated chromosome rearrangements after divergence from a common ancestor of the genus *Brassica*.

In summary, the DNA repeats presented in our study are core constituents of *Brassica* heterochromatin. While the repeats were found to be generally conserved within *Brassica*, some species-specific divergence was observed. These differences include the absence of CentBr in the *Brassica* B-genome, the divergence of CentBr1 and CentBr2 in *Brassica* species, the absence of pBgt in the *Brassica* C-genome, and the *Brassica* lineage-specific divergence of the 350-bp subtelomeric tandem repeat. Our results provide a framework for the detailed analysis of heterochromatin during the whole-genome sequencing of *Brassica* and related species. The chromosomal locations and abundance of repetitive heterochromatin in *Brassica* species and *R. sativus* provide insight into chromosome evolution in these species. Overall, these results demonstrate that *Brassica* species have highly dynamic genomes that reflect the great phenotypic diversity and plasticity of the Brassicaceae.

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2010.12.002.

## Acknowledgments

This work was supported by grants for the Technology Development Program for Agriculture and Forestry (Grant no. 607002-05), Ministry of Food, Agriculture, Forestry, and Fisheries, Republic of Korea. The authors thank the late Dr. Prikshit Plaha for reviewing the manuscript.

## References

[1] J.S. Heslop-Harrison, Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes, Plant Cell 12 (2000) 617–636.
[2] J. Messing, et al., Sequence composition and genome organization of maize, Proc. Natl Acad. Sci. USA 101 (2004) 14349–14354.
[3] W. Li, P. Zhang, J.P. Fellers, B. Friebe, B.S. Gill, Sequence composition, organization, and evolution of the core Triticeae genome, Plant J. 40 (2004) 500–511.
[4] Identification and distribution of seven classes of middle-repetitive DNA in the *Arabidopsis thaliana* genome, Nucleic Acids Res. 24 (1996) 3017–3022.
[5] S. Henikoff, K. Ahmad, H.S. Malik, The centormere paradox: stable inheritance with rapidly evolving DNA, Science 293 (2001) 1098–1102.
[6] H.R. Lee, et al., Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species, Proc. Natl Acad. Sci. USA 102 (2005) 11793–11798.
[7] K. Nagaki, et al., Chromatin immunoprecipitation reveals that the 180 bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres, Genetics 163 (2003) 1221–1225.
[8] J.F. Wendel, S.R. Wessler, Retrotransposon-mediated genome evolution on a local ecological scale, Proc. Natl Acad. Sci. USA 97 (2000) 6250–6252.
[9] N. Fedoroff, Transposons and genome evolution in plants, Proc. Natl Acad. Sci. USA 97 (2000) 7002–7007.
[10] N. U, Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization, Jpn. J. Bot. 7 (1935) 389–452.
[11] Y.W. Yang, K.N. Lai, P.Y. Tai, W.H. Li, Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages, J. Mol. Evol. 48 (1999) 597–604.
[12] F. Cheung, et al., Comparative analysis between homoeologous genome segments of *Brassica napus* and its progenitor species reveals extensive sequence-level divergence, Plant Cell 21 (2006) 1912–1918.
[13] M.A. Lysak, M.A. Koch, A. Pecinka, I. Schubert, Chromosome triplication found across the tribe *Brassiceae*, Genome Res. 15 (2005) 516–525.
[14] J. Maluszynska, J.S. Heslop-Harrison, Physical mapping of rDNA loci in *Brassica* species, Genome 36 (1993) 774–781.
[15] E. Harrison, J.S. Heslop-Harrison, Centromeric repetitive DNA sequences in the genus *Brassica*, Theor. Appl. Genet. 90 (1995) 157–165.
[16] D.H. Koo, P. Plaha, Y.P. Lim, Y. Hur, J.W. Bang, A high-resolution karyotype of *Brassica rapa* ssp. *pekinensis* revealed by pachytene analysis and multicolor fluorescence *in situ* hybridization, Theor. Appl. Genet. 109 (2004) 1346–1352.
[17] K.B. Lim, et al., Characterization of rDNAs and tandem repeats in the heterochromatin of *Brassica rapa*, Mol. Cells 19 (2005) 436–444.
[18] K.B. Lim, et al., Characterization of the centromere and peri-centromere retrotransposons in *Brassica rapa* and their distribution in related *Brassica* species, Plant J. 49 (2007) 173–183.
[19] C.J. Schelfhout, R. Snowdon, W.A. Cowling, J.M. Wroth, A PCR based B-genome-specific marker in *Brassica* species, Theor. Appl. Genet. 109 (2004) 917–921.
[20] R. Hasterok, T. Ksiazczyk, E. Wolny, J. Maluszynska, FISH and GISH analysis of *Brassica* genomes, Acta Biol. Crac. 47 (2005) 185–192.
[21] K. Alix, C.D. Ryder, J. Moore, G.J. King, J.S. Heslop-Harrison, The genomic organization of retrotransposons in *Brassica oleracea*, Plant Mol. Biol. 59 (2005) 839–851.
[22] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular cloning: a laboratory manual (2nd ed), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
[23] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[24] A.V. Cox, et al., Comparison of plant telomere locations using a PCR-generated synthetic probe, Ann. Bot. 72 (1993) 239–247.
[25] S. Jackson, M.L. Wang, H.M. Goodman, J. Jiang, Application of fiber-FISH in physical mapping of *Arabidopsis thaliana*, Genome 41 (1998) 566–572.

[26] B. Ewing, P. Green, Base-calling of automated sequencer traces using *phred*. II. Error probabilities, Genome Res. 8 (1998) 186–194.
[27] X. Xia, P.S. Rocha, G. Selvaraj, H. Bertrand, Structure and evolution of a highly repetitive DNA sequence from *Brassica napus*, Plant Mol. Biol. 21 (1993) 213–224.
[28] X. Xia, P.S. Rocha, G. Selvaraj, H. Bertrand, Genomic organization of the repetitive DNA in *Brassica juncea*, Plant Mol. Biol. 26 (1994) 817–832.
[29] F. Grellet, D. Delcasso, F. Panabieres, M. Delseny, Organization and evolution of a higher plant alphoid-like satellite DNA sequence, J. Mol. Biol. 187 (1986) 495–507.
[30] S. Harbinder, M. Lakshmikumaran, A repetitive sequence from *Diplotaxis erucoides* is highly homologous to that of *Brassica campestris* and *B. oleracea*, Plant Mol. Biol. 15 (1990) 155–156.
[31] J.S. Heslop-Harrison, M. Murata, Y. Ogura, T. Schwarzacher, F. Motoyoshi, Polymorphisms and genomic organization of repetitive DNA from centromeric regions of Arabidopsis chromosomes, Plant Cell 11 (1999) 31–42.
[32] A. Kawabe, S. Nasuda, Polymorphic chromosomal specificity of centromere satellite families in *Arabidopsis halleri* ssp. *gemmifera*, Genetica 126 (2006) 335–342.
[33] A. Berr, et al., Chromosome arrangement and nuclear architecture but not centromeric sequences are conserved between *Arabidopsis thaliana* and *Arabidopsis lyrata*, Plant J. 48 (2006) 771–783.
[34] A. Kawabe, S. Nasuda, Chromosome-specific satellite sequences in *Turritis glabra*, Genes Genet. Syst. 81 (2006) 287–290.
[35] S.E. Hall, S. Luo, A.E. Hall, D. Preuss, Differential rates of local and global homogenization in centromere satellites from Arabidopsis relatives, Genetics 170 (2005) 1913–1927.
[36] C.D. Bailey, et al., Toward a global phylogeny of the Brassicaceae, Mol. Biol. Evol. 23 (2006) 2142–2160.
[37] E.J. Richards, F.M. Ausubel, Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*, Cell 53 (1988) 127–136.
[38] J.F. Elder, B.J. Turner, Concerted evolution at the population level: pupfish *Hin*dIII satellite DNA sequences, Proc. Natl Acad. Sci. USA 91 (1994) 994–998.
[39] Z. Cheng, et al., Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon, Plant Cell 14 (2002) 1691–1704.
[40] G.P. Copenhaver, et al., Genetic definition and sequence analysis of *Arabidopsis* centromeres, Science 286 (1999) 2468–2474.
[41] W. Jin, et al., Maize centromeres: organization and functional adaptation in the genetic background of oat, Plant Cell 16 (2004) 571–581.
[42] H. Ohtsubo, E. Ohtsubo, Involvement of transposition in dispersion of tandem repeat sequences (TrsA) in rice genomes, Mol. Gen. Genet. 245 (1994) 449–455.
[43] P. Zhang, B. Friebe, B.S. Gill, Variation in the distribution of a genome-specific DNA sequence on chromosomes reveals evolutionary relationships in the *Triticum* and *Aegilops* complex, Plant Syst. Evol. 235 (2002) 169–179.
[44] M.W. Ganal, N.L. Lapitan, S.D. Tanksley, Macrostructure of the tomato telomeres, Plant Cell 3 (1991) 87–94.
[45] H. Weiss, H. Scherthan, *Aloe* spp. — plants with vertebrate-like telomeric sequences, Chromosome Res. 10 (2002) 155–164.
[46] E. S korová, et al., A.R. Leitch, Minisatellite telomeres occur in the family Alliaceae but are lost in *Allium*, Am. J. Bot. 93 (2006) 814–823.
[47] J.W. Ijdo, A. Baldini, D.C. Ward, S.T. Reeders, R.A. Wells, Origin of human chromosome 2: an ancestral telomere-telomere fusion, Proc. Natl Acad. Sci. USA 88 (1991) 9051–9055.
[48] H. Biessmann, J.M. Mason, Telomeric repeat sequences, Chromosoma 103 (1994) 154–161.